Loading dataset...

Dataset loaded: 28766 samples, 20 features

Class distribution: {0: 18514, 1: 10252}

Initializing feature selection techniques...

==================================================

Processing 1. Chi-Square

==================================================

Selecting features using 1. Chi-Square...

Top 10 features selected by 1. Chi-Square:

1. Breathing Problem: 9964.0393

2. Sore throat: 13201.1649

3. Heart Disease: 3552.0005

4. Diabetes: 1929.1672

5. Hyper Tension: 3089.5719

6. Gastrointestinal : 1693.0652

7. Abroad travel: 11288.4346

8. Contact with COVID Patient: 8194.1488

9. Attended Large Gathering: 9140.5020

10. Family working in Public Exposed Places: 7781.7278

Training CNN with features selected by 1. Chi-Square

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 5ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 6ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 5ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **2s** 5ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **2s** 7ms/step

Results for 1. Chi-Square:

Average accuracy: 0.9833 ± 0.0040

Average precision: 0.9941 ± 0.0036

Average recall: 0.9592 ± 0.0123

Average f1: 0.9763 ± 0.0055

Average auc: 0.9987 ± 0.0003

==================================================

Processing 2. Mutual Information

==================================================

Selecting features using 2. Mutual Information...

Top 10 features selected by 2. Mutual Information:

1. Breathing Problem: 0.3088

2. Sore throat: 0.4123

3. Heart Disease: 0.0913

4. Diabetes: 0.1108

5. Hyper Tension: 0.0735

6. Gastrointestinal : 0.0817

7. Abroad travel: 0.2871

8. Contact with COVID Patient: 0.1847

9. Attended Large Gathering: 0.2381

10. Family working in Public Exposed Places: 0.2183

Training CNN with features selected by 2. Mutual Information

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 6ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 5ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 5ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Results for 2. Mutual Information:

Average accuracy: 0.9845 ± 0.0023

Average precision: 0.9956 ± 0.0035

Average recall: 0.9608 ± 0.0061

Average f1: 0.9778 ± 0.0029

Average auc: 0.9986 ± 0.0003

==================================================

Processing 3. Recursive Feature Elimination

==================================================

Selecting features using 3. Recursive Feature Elimination...

Top 10 features selected by 3. Recursive Feature Elimination:

1. Breathing Problem: 1.0000

2. Fever: 1.0000

3. Dry Cough: 1.0000

4. Sore throat: 1.0000

5. Hyper Tension: 1.0000

6. Abroad travel: 1.0000

7. Contact with COVID Patient: 1.0000

8. Attended Large Gathering: 1.0000

9. Visited Public Exposed Places: 1.0000

10. Family working in Public Exposed Places: 1.0000

Training CNN with features selected by 3. Recursive Feature Elimination

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Results for 3. Recursive Feature Elimination:

Average accuracy: 0.9787 ± 0.0013

Average precision: 0.9789 ± 0.0063

Average recall: 0.9609 ± 0.0064

Average f1: 0.9698 ± 0.0015

Average auc: 0.9979 ± 0.0004

==================================================

Processing 4. Lasso

==================================================

Selecting features using 4. Lasso...

Top 10 features selected by 4. Lasso:

1. Breathing Problem: 1.0000

2. Sore throat: 1.0000

3. Fatigue : 1.0000

4. Abroad travel: 1.0000

5. Contact with COVID Patient: 1.0000

6. Attended Large Gathering: 1.0000

7. Visited Public Exposed Places: 1.0000

8. Family working in Public Exposed Places: 1.0000

9. Fever: 0.0000

10. Dry Cough: 0.0000

Training CNN with features selected by 4. Lasso

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 3ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Results for 4. Lasso:

Average accuracy: 0.9780 ± 0.0032

Average precision: 0.9728 ± 0.0135

Average recall: 0.9657 ± 0.0081

Average f1: 0.9691 ± 0.0038

Average auc: 0.9979 ± 0.0003

==================================================

Processing 5. Random Forest Importance

==================================================

Selecting features using 5. Random Forest Importance...

Top 10 features selected by 5. Random Forest Importance:

1. Breathing Problem: 0.1560

2. Sore throat: 0.2333

3. Abroad travel: 0.1637

4. Contact with COVID Patient: 0.0630

5. Attended Large Gathering: 0.1179

6. Family working in Public Exposed Places: 0.0940

7. Fever: 0.0000

8. Dry Cough: 0.0000

9. Running Nose: 0.0000

10. Asthma: 0.0000

Training CNN with features selected by 5. Random Forest Importance

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 3ms/step

Results for 5. Random Forest Importance:

Average accuracy: 0.9822 ± 0.0016

Average precision: 0.9845 ± 0.0080

Average recall: 0.9654 ± 0.0057

Average f1: 0.9748 ± 0.0021

Average auc: 0.9981 ± 0.0003

==================================================

Processing 6. Boruta

==================================================

Selecting features using 6. Boruta...

Top 10 features selected by 6. Boruta:

1. Breathing Problem: 1.0000

2. Sore throat: 1.0000

3. Running Nose: 1.0000

4. Asthma: 1.0000

5. Chronic Lung Disease: 1.0000

6. Headache: 1.0000

7. Heart Disease: 1.0000

8. Diabetes: 1.0000

9. Hyper Tension: 1.0000

10. Fatigue : 1.0000

Training CNN with features selected by 6. Boruta

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Results for 6. Boruta:

Average accuracy: 0.9724 ± 0.0111

Average precision: 0.9707 ± 0.0297

Average recall: 0.9523 ± 0.0027

Average f1: 0.9612 ± 0.0149

Average auc: 0.9952 ± 0.0027

Loading dataset...

Dataset loaded: 28766 samples, 20 features

Class distribution: {0: 18514, 1: 10252}

Initializing feature selection techniques...

==================================================

Processing 7. Correlation-based

==================================================

Selecting features using 7. Correlation-based...

Top 10 features selected by 7. Correlation-based:

1. Sore throat: 0.8628

2. Breathing Problem: 0.7620

3. Attended Large Gathering: 0.6493

4. Family working in Public Exposed Places: 0.6421

5. Abroad travel: 0.6099

6. Contact with COVID Patient: 0.5177

7. Diabetes: 0.4520

8. Heart Disease: 0.4011

9. Gastrointestinal : 0.3945

10. Hyper Tension: 0.3857

Training CNN with features selected by 7. Correlation-based

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Results for 7. Correlation-based:

Average accuracy: 0.9837 ± 0.0035

Average precision: 0.9884 ± 0.0098

Average recall: 0.9659 ± 0.0095

Average f1: 0.9769 ± 0.0046

Average auc: 0.9986 ± 0.0004

==================================================

Processing 8. Sequential Forward Selection

==================================================

Selecting features using 8. Sequential Forward Selection...

Top 10 features selected by 8. Sequential Forward Selection:

1. Fever: 1.0000

2. Dry Cough: 1.0000

3. Sore throat: 1.0000

4. Running Nose: 1.0000

5. Asthma: 1.0000

6. Chronic Lung Disease: 1.0000

7. Abroad travel: 1.0000

8. Attended Large Gathering: 1.0000

9. Wearing Masks: 1.0000

10. Sanitization from Market: 1.0000

Training CNN with features selected by 8. Sequential Forward Selection

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Results for 8. Sequential Forward Selection:

Average accuracy: 0.9586 ± 0.0044

Average precision: 0.9814 ± 0.0122

Average recall: 0.9015 ± 0.0190

Average f1: 0.9395 ± 0.0064

Average auc: 0.9912 ± 0.0006

==================================================

Processing 9. XGBoost Importance

==================================================

Selecting features using 9. XGBoost Importance...

Top 10 features selected by 9. XGBoost Importance:

1. Sore throat: 0.5873

2. Abroad travel: 0.2135

3. Attended Large Gathering: 0.0666

4. Breathing Problem: 0.0000

5. Fever: 0.0000

6. Dry Cough: 0.0000

7. Running Nose: 0.0000

8. Asthma: 0.0000

9. Chronic Lung Disease: 0.0000

10. Headache: 0.0000

Training CNN with features selected by 9. XGBoost Importance

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Results for 9. XGBoost Importance:

Average accuracy: 0.9761 ± 0.0046

Average precision: 0.9819 ± 0.0080

Average recall: 0.9508 ± 0.0163

Average f1: 0.9659 ± 0.0067

Average auc: 0.9976 ± 0.0006

==================================================

Processing 10. LightGBM Importance

==================================================

Selecting features using 10. LightGBM Importance...

[LightGBM] [Warning] Found whitespace in feature\_names, replace with underlines

[LightGBM] [Info] Number of positive: 10252, number of negative: 18514

[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.002154 seconds.

You can set `force\_row\_wise=true` to remove the overhead.

And if memory is not enough, you can set `force\_col\_wise=true`.

[LightGBM] [Info] Total Bins 32

[LightGBM] [Info] Number of data points in the train set: 28766, number of used features: 16

[LightGBM] [Info] [binary:BoostFromScore]: pavg=0.356393 -> initscore=-0.591054

[LightGBM] [Info] Start training from score -0.591054

Top 10 features selected by 10. LightGBM Importance:

1. Breathing Problem: 266.0000

2. Sore throat: 212.0000

3. Asthma: 220.0000

4. Chronic Lung Disease: 240.0000

5. Heart Disease: 207.0000

6. Hyper Tension: 207.0000

7. Abroad travel: 169.0000

8. Contact with COVID Patient: 216.0000

9. Attended Large Gathering: 183.0000

10. Family working in Public Exposed Places: 199.0000

Training CNN with features selected by 10. LightGBM Importance

Training fold 1/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **0s** 2ms/step

Training fold 2/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 3/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 4/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Training fold 5/5...

**180/180** ━━━━━━━━━━━━━━━━━━━━ **1s** 2ms/step

Results for 10. LightGBM Importance:

Average accuracy: 0.9858 ± 0.0021

Average precision: 0.9890 ± 0.0064

Average recall: 0.9711 ± 0.0035

Average f1: 0.9800 ± 0.0026

Average auc: 0.9986 ± 0.0003

Comparison chart saved as 'feature\_selection\_comparison.png'

Heatmap saved as 'feature\_selection\_heatmap.png'

==================================================

FINAL SUMMARY

==================================================

Techniques ranked by accuracy:

1. 10. LightGBM Importance: 0.9858

2. 2. Mutual Information: 0.9845

3. 7. Correlation-based: 0.9837

4. 1. Chi-Square: 0.9833

5. 5. Random Forest Importance: 0.9822

6. 3. Recursive Feature Elimination: 0.9787

7. 4. Lasso: 0.9780

8. 9. XGBoost Importance: 0.9761

9. 6. Boruta: 0.9724

10. 8. Sequential Forward Selection: 0.9586

Best performing technique: 10. LightGBM Importance

Top 10 features selected by 10. LightGBM Importance:

1. Breathing Problem

2. Sore throat

3. Asthma

4. Chronic Lung Disease

5. Heart Disease

6. Hyper Tension

7. Abroad travel

8. Contact with COVID Patient

9. Attended Large Gathering

10. Family working in Public Exposed Places